National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

A Framework to Enable Interpretation of the Data

J. Steven Hughes
Principal Computer Scientist
Jet Propulsion Laboratory

ESIP Winter Meeting 2017
Bethesda, Maryland

Use Cases

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?

- What coordinate system was used for the HiRISE RDR images?

- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
  - *the calibration files used to calibrate this image*
  - *published instrument design documents.*

https://en.wikipedia.org/wiki/Mars_Reconnaissance_Orbiter

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Terms, Definitions, and Relationships

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?

The Raw Data

- What coordinate system was used for the HiRISE RDR images?

- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
  - *the calibration files used to calibrate this image*
  - *published instrument design documents.*

Terms, Definitions, and Relationships

What are these things?

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?

- What coordinate system was used for the HiRISE RDR images?

- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
  - *the calibration files used to calibrate this image*
  - *published instrument design documents.*

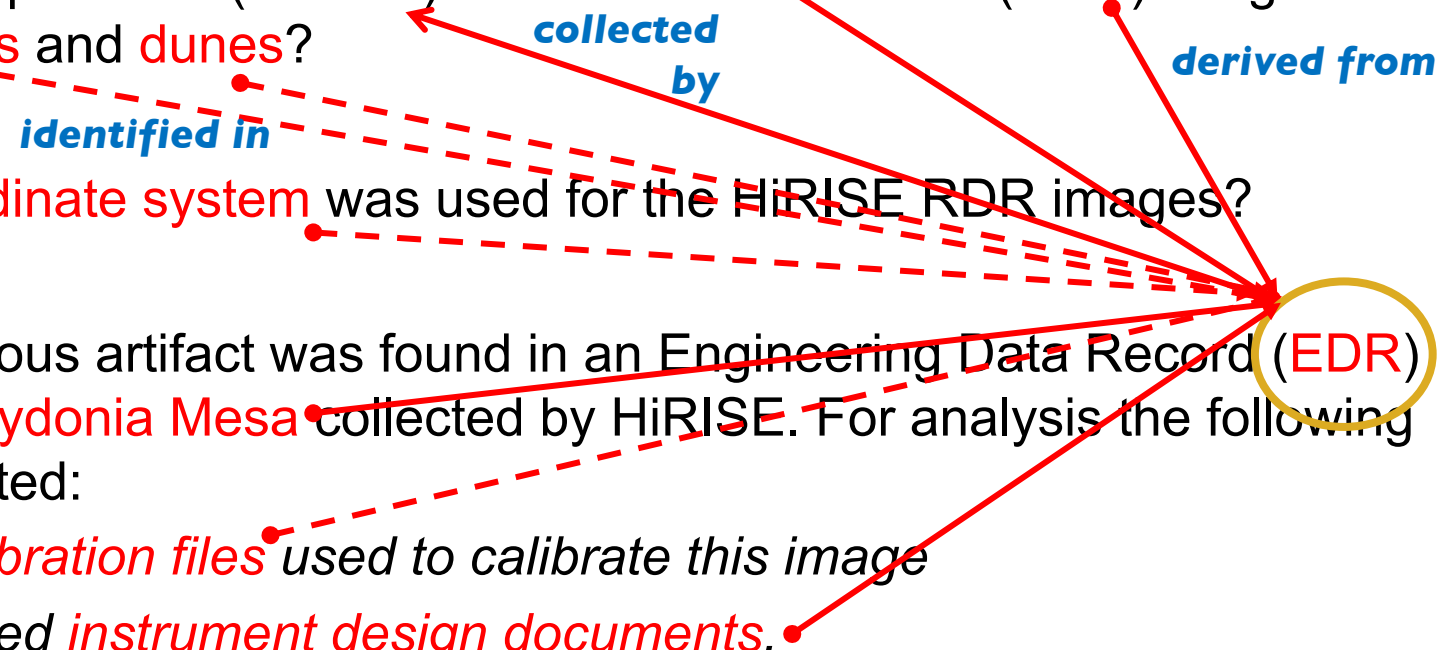
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Terms, Definitions, and Relationships

How are these things related to the data?

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?
- What coordinate system was used for the HiRISE RDR images?
- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
  - *the calibration files used to calibrate this image*
  - *published instrument design documents.*



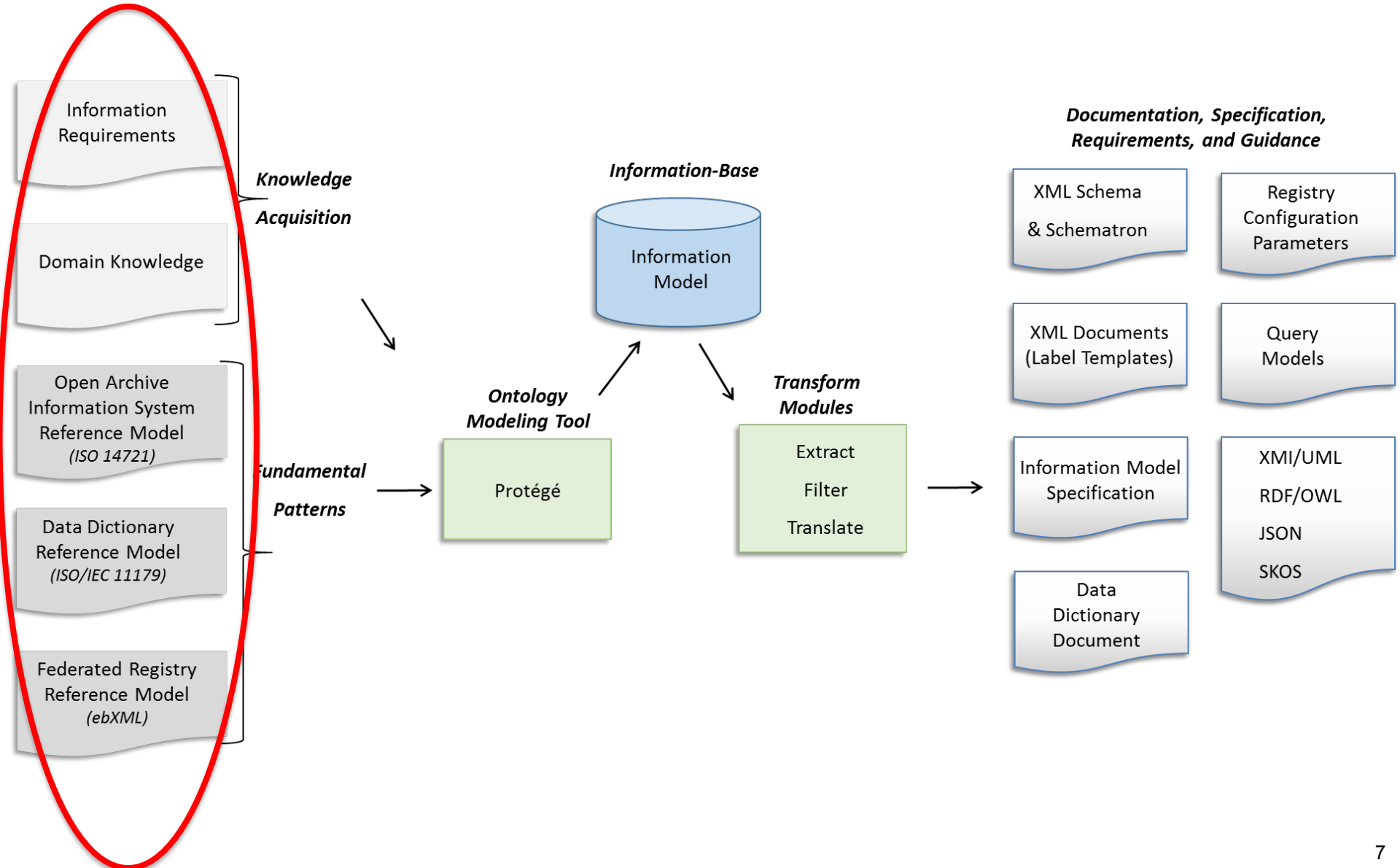
The Information Model (IM)

- The Information Model provides the information requirements for the system
  - *Defines the terms in the community and their relationships*
  - *Improves consistency and interoperability and reduces ambiguity*
- Establishes an overarching federated governance model for the metadata
  - *Provides common, discipline, and local governance*
  - *Localizes changes and allows extensions as the community evolves*
  - *Allows for model independence*
- Establishes the “corner-stone” of the “information model-driven” driven design paradigm
  - *Allows the system to be configured by and to respond to the information model (information requirements)*
  - *Enables agile development*
  - *Handles diversity*
    - Accommodates new instruments, observation types, data, …
  - *Reduce the impact of changes on the system*

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Components of the Information Model



The Input to the Model Consists of Two Parts

- Common Foundational Principles
  - *Open Archival Information System Reference Model (OAIS-RM)*
  - *Data Dictionary Reference Model*
  - *Federated Reference Model*

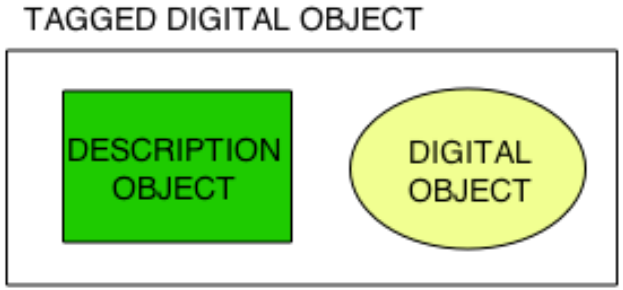
- Community-Specific Input
  - *Information Requirements*
  - *Domain Knowledge*

National Aeronautics and Space Administration

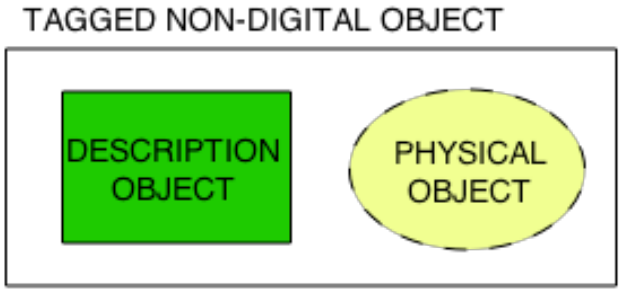
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Open Information Archive System (OAIS) Reference Model

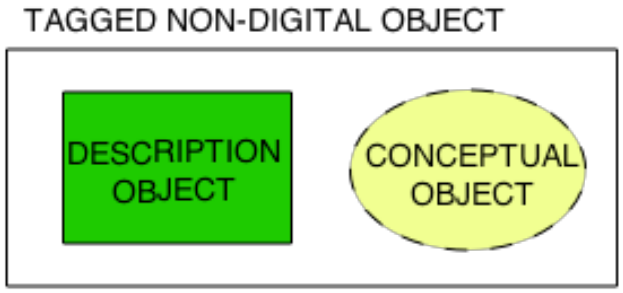
Information Object Model[1]



- **digital object**: An object which is real data — for example, a binary image of a redwood tree.



- **physical object**: An object which is physical or tangible – for example the planet Saturn and the Venus Express magnetometer.



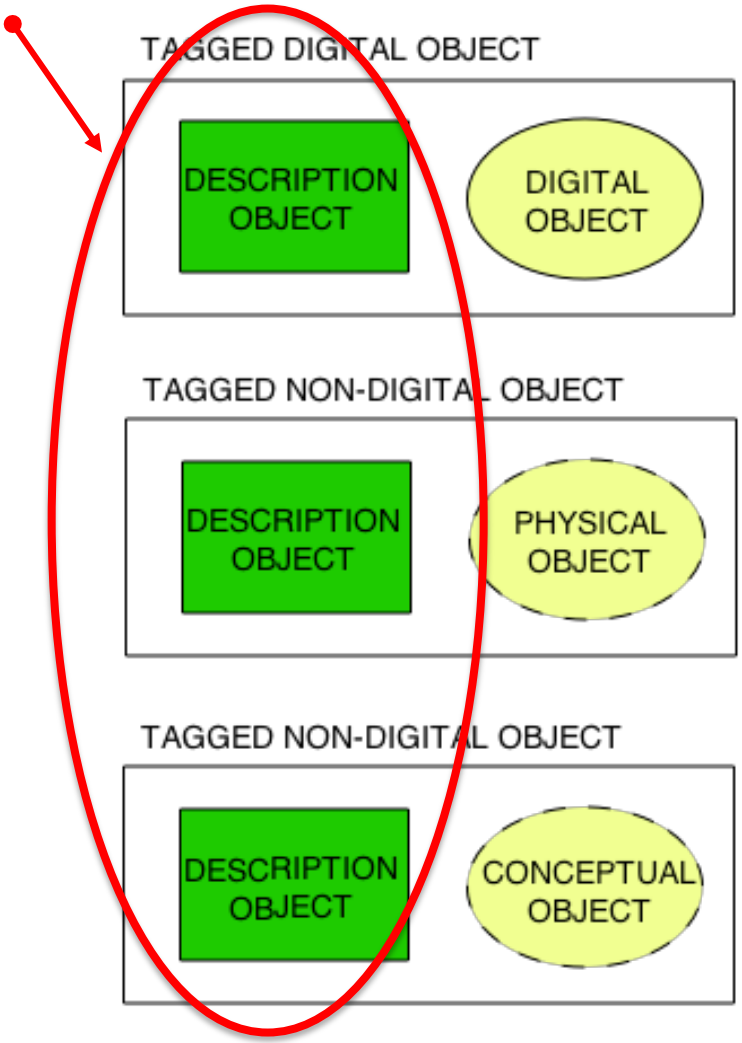
- **conceptual object**: An object which is intangible – for example the Cassini mission and NASA's strategic plan for solar system exploration.

[1]Open Archival Information System (OAIS) Reference Model - ISO 14721:2003

Data Management[1]

Information Object Model[1]

Description Object Management



- **digital object**: An object which is real data — for example, a binary image of a redwood tree.
- **physical object**: An object which is physical or tangible – for example the planet Saturn and the Venus Express magnetometer.
- **conceptual object**: An object which is intangible – for example the Cassini mission and NASA's strategic plan for solar system exploration.

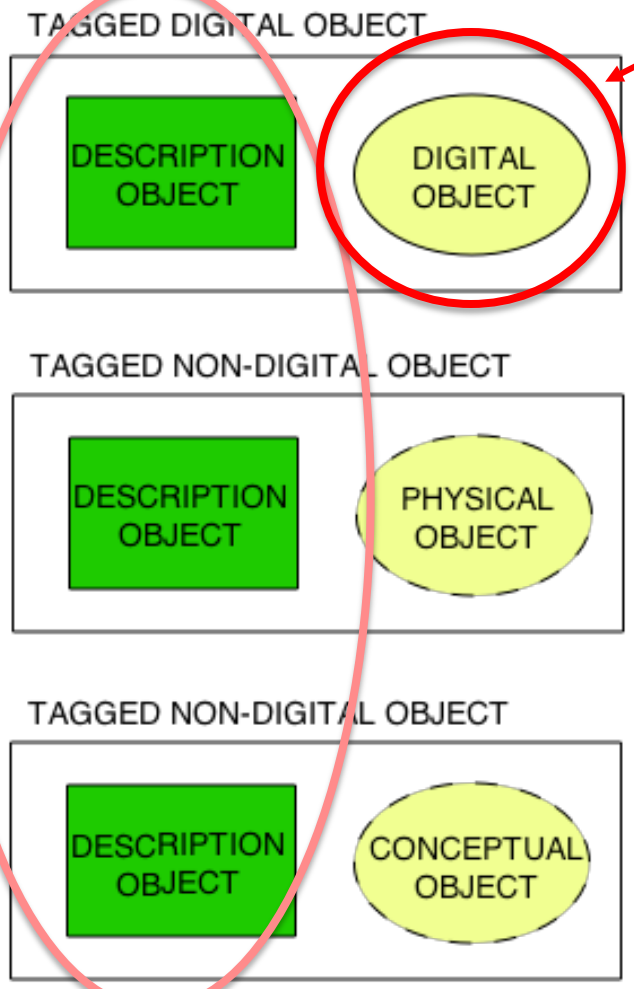
[1]Open Archival Information System (OAIS) Reference Model - ISO 14721:2003

Data Object Management

Information Object Model[1]

Description Object Management

Data Object Management



- **digital object**: An object which is real data — for example, a binary image of a redwood tree.
- **physical object**: An object which is physical or tangible – for example the planet Saturn and the Venus Express magnetometer.
- **conceptual object**: An object which is intangible – for example the Cassini mission and NASA's strategic plan for solar system exploration.

[1]Open Archival Information System (OAIS) Reference Model - ISO 14721:2003

Information Categories[1]

- **Identification**
  - *Identification information provides a unique and immutable identifier for any information object that is to be discovered and accessed.*
- **Representation/Format**
  - *Representation information allows a data object to be interpreted. This includes describing the data format.*
- **Integrity (Fixity)**
  - *Integrity information ensures the information object has not been unintentionally altered.*
- **Provenance**
  - *Provence Information provides the history of the data and is essential for authenticity. It must include the provider.*
- **Context**
  - *Context information provides additional information that describes the environment in which the data object was created. For example, context information may describe instruments or light sources.*
- **Reference**
  - *Reference information allows the information objects to be referenced. Identification information is a subset of Reference Information.*
- **Access Rights**
  - *Access Rights information identifies the access restrictions pertaining to the data, including the legal framework, licensing terms, and access control; provider provided access and distribution conditions, and specifications for rights enforcement measures.*

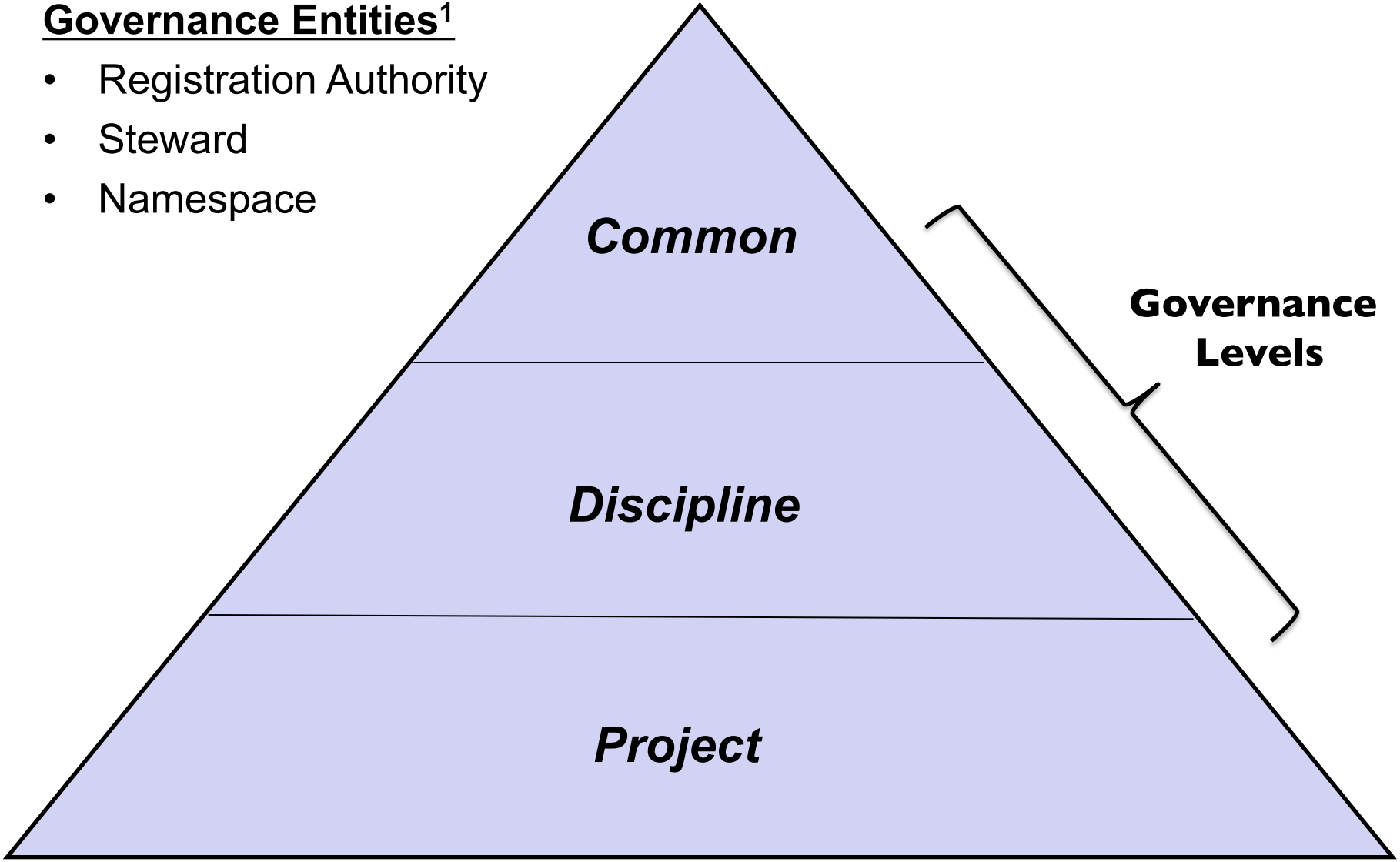
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data Dictionary Reference Model[2]

Governance Entities[1]

- Registration Authority
- Steward
- Namespace



[2]ISO/IEC 11179:2003 - Volume 3 – Metadata Registry Specification

Registry Reference Model

- ebXML (Electronic Business XML) - Standardizes the secure exchange of data

- Defines key properties of a federated registry
  - *registry database schema*
  - *registry object (generic)*
    - Extensions: products, granules, etc
    - First class objects
      - *digital, physical, and conceptual*
  - *core attributes*
    - identification (e.g. <logical_identifier>)
    - versioning (e.g. <version_identifier>)
  - *common registry services*
    - tracking/locate/retrieval

Information Requirements

- The community's Requirements and Policies set the foundation for the *information requirements*.

1. The System will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting the institutions goals in science exploration.

    1.4 Archiving Standards: The system will have **archiving standards** for science data

        1.4.1 The system will **define a standard** for organizing, formatting, and documenting science data

        1.4.2 The system will maintain a **dictionary of terms, values, and relationships** for standardized description of science data

        1.4.3 The system will define a **standard grammar** for describing science data

        1.4.4 The system will **establish minimum content requirements** for a data set (primary and ancillary data)

        1.4.5 The system will, for each mission or other major data provider, produce a list of the **minimum components required for archival data**

        1.4.6 The system will **develop, publish and implement a process for managing changes** to the archive standards

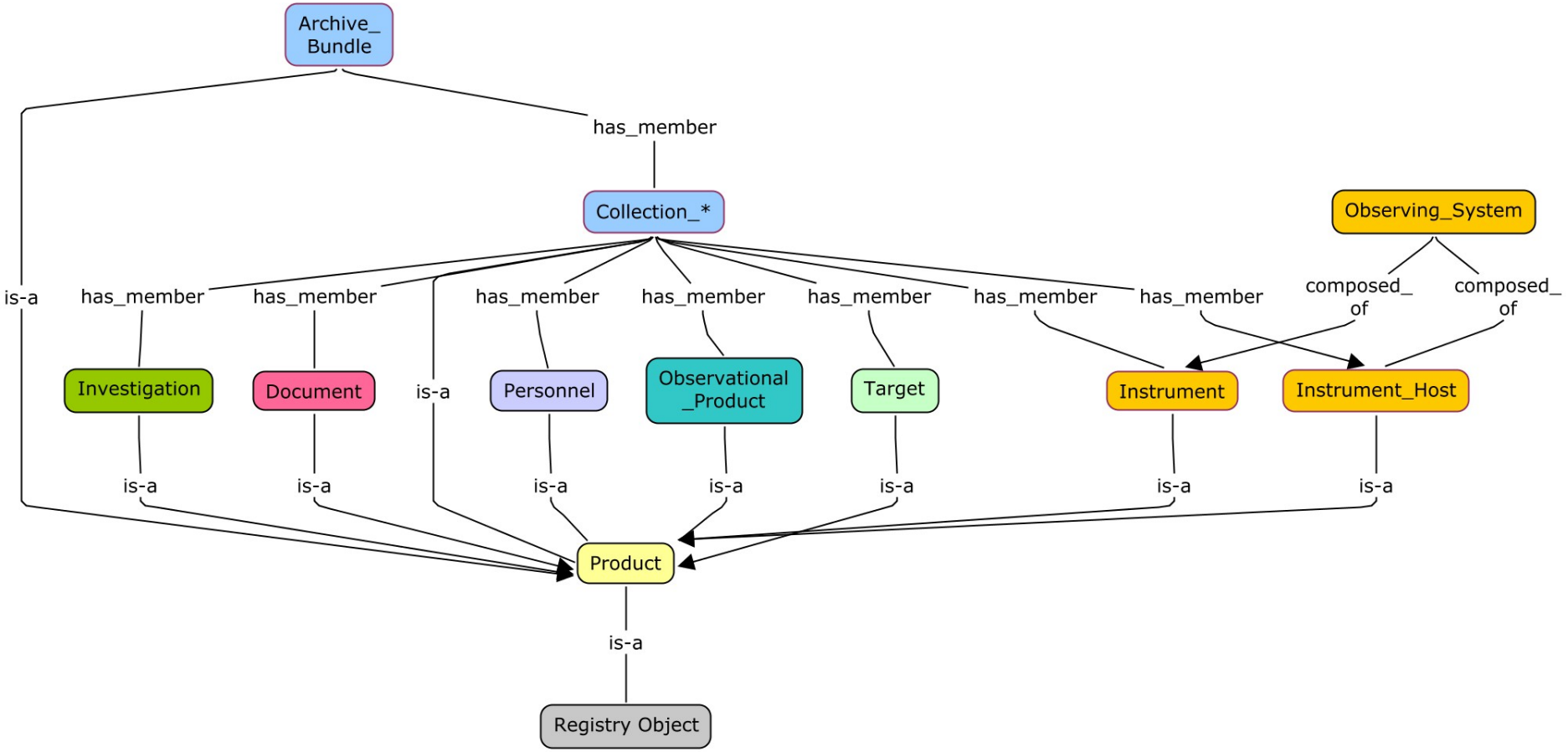
        1.4.7 The system will keep **abreast of new developments** in archiving standards

Domain Knowledge

- The experiential (first hand) information about things of interest known by experts in the domain.

- Information about the “things” that should be collected and associated with the data to make and keep it useful.
  - *The data and their structures (representation information)*
  - *The context within which the data was used and collected*
    - Investigations/Missions/Campaigns
    - Observing Systems/Instruments
    - Personnel
    - Data collection targets of interest

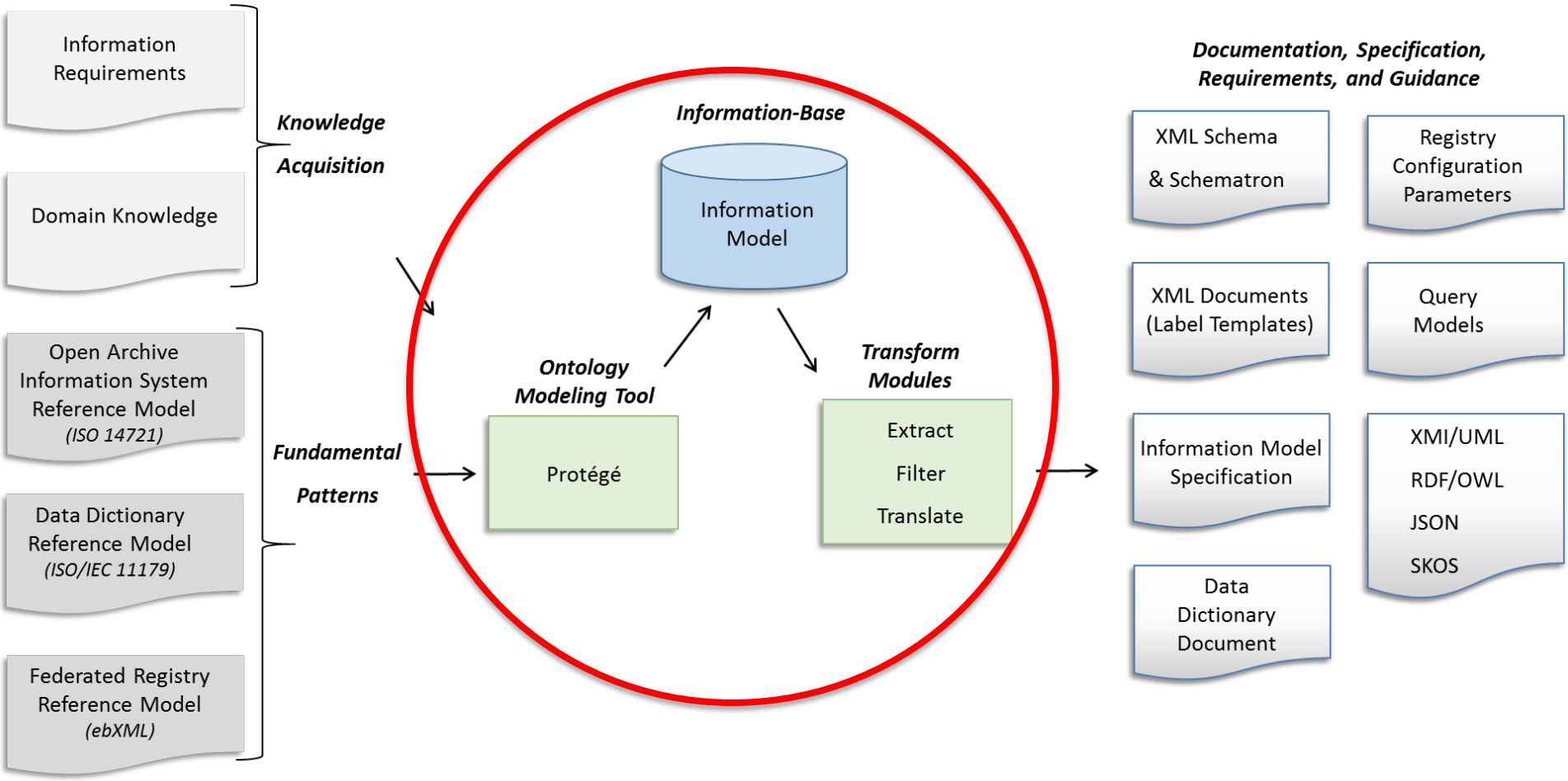
Domain Knowledge



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model Management



The Information Model Database

- All “things of interest” are defined in an object modeling tool.
  - *All objects and their attributes and relationships.*
  - *Typically an ontology modeling tool is used*
    - Necessary but not necessarily sufficient
- A master database is created by merging the object models and the data dictionary.
- The contents of the master database is filtered and written to system files in various formats.
  - *Used by the data providers, registry, harvester, search engine, validator and other system tools and services.*

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data Dictionary Schema

Data Element

- *Name*
- *Submitter, Steward*
- *Definition*
- *Namespace*
- *Source of definition*
- *Change log*
- *Version*
- *Concept*
- *Alternate Names*
- *Definition in different natural languages*
- *Classification*
- *Unit of measurement*
- *Effective Dates*

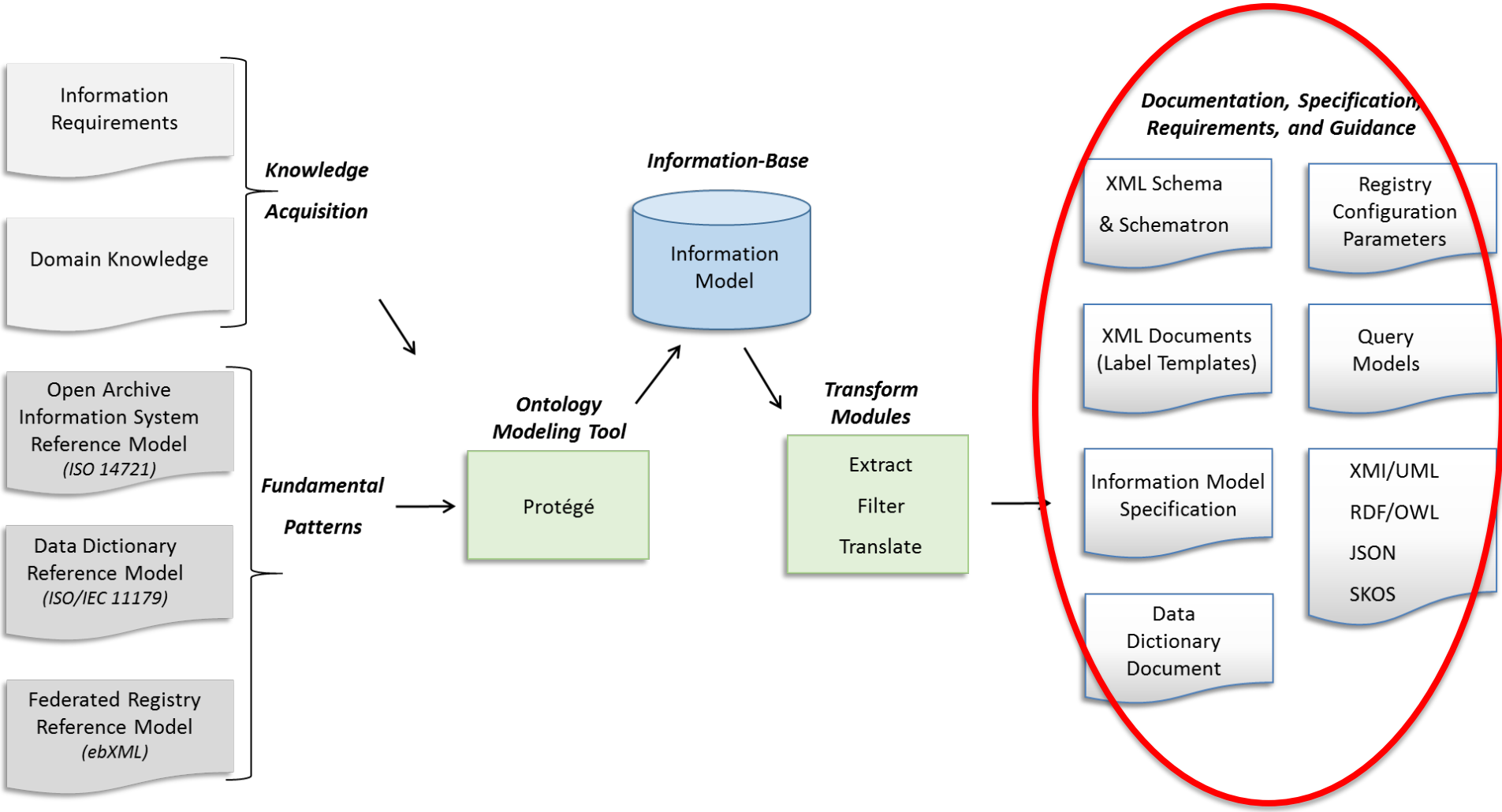
Value Domain

- Permissible Value
- Value Meaning
- Submitter, Steward
- Definition
- Cardinality
- Source of definition
- Change log
- Version
- Concept
- Character Set
- Representation
- Minimum and Maximum Value
- Minimum and Maximum Length
- Alternate encodings
- Effective Dates

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model Output



XML Schema and Schematron Files

```
<xs:complexType name="Array">
  <xs:annotation>
    <xs:documentation>The Array class defines a homogeneous N-dimensional array of scalars. …
  </xs:annotation>
  <xs:complexContent>
    <xs:extension base="pds:Byte_Stream">
      <xs:sequence>
        <xs:element name="offset" type="pds:offset" minOccurs="1" maxOccurs="1"> </xs:element>
        <xs:element name="axes" type="pds:axes" minOccurs="1" maxOccurs="1"> </xs:element>
        <xs:element name="axis_index_order" type="pds:axis_index_order" minOccurs="1“ …
        <xs:element name="description" type="pds:description" minOccurs="0" maxOccurs="1"> …
        <xs:element name="Element_Array" type="pds:Element_Array" minOccurs="1“  …
        <xs:element name="Axis_Array" type="pds:Axis_Array" minOccurs="1“ …
          …


  <sch:pattern>
    <sch:rule context="pds:Array/pds:axis_index_order">
      <sch:assert test=". = ('Last Index Fastest')">
        The attribute pds:axis_index_order must be equal to the value 'Last Index Fastest'.</sch:assert>
```

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Product Label Template

Identification_Area
- Logical_Identifier
- Version_Id

Observation_Area
- Time_Coordinates
- Primary_Result_Summary
- Investigation_Area
- Observing_System
- Target_Identification

Discipline_Area
Mission Area

Reference_List
- Internal_Reference
- External_Reference

File_Area_Observational
- File
  - Header
  - Array_2d_Image

The Framework at Work

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?
  - *Spacecraft, instruments, digital images, documentation, and calibration files are all defined in the **information model**.*
  - ***Labeled objects** are created, ingested, and related in the registry.*
  - *All **registered objects** are considered first-class.*
  - *The Imaging Atlas (catalog) uses semantics from the information model and harvested metadata from labeled objects to provide **field-** and **facet-based** search.*
  - *Image content **annotations** were generated using a visual salience **landmark** detector plus a deep learning neural network classifier.*
  - *Enables Image Atlas to provide **content-based** search for HiRISE RDR images of Mars*

Semantics at Work

- What coordinate system was used for the HiRISE RDR images?
  - *The coordinate system used is **planetocentric** latitude and east positive longitude direction*
  - *Coordinate systems are defined as a class in a **discipline** level cartography model*

- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
  - *the calibration files used to calibrate this image*
  - *published instrument design documents.*

  - ***Documents** are either referenced as registered objects or via bibliographic citations / DOIs.*
  - ***Features** are classified and defined in a feature catalog.*

Thank You!

PDS4 Documents
https://pds.jpl.nasa.gov/pds4/doc/index.shtml

Backup